Analysis of the Variance Threshold of Kay's Weighted Linear Predictor Frequency Estimator

Vaughan Clarkson, *Student Member, IEEE,* Peter J. Kootsookos, *Member, IEEE,* and Barry G. Quinn, *Member, IEEE*

***Abstract*— A theoretical approximation for the variance of Kay's weighted linear predictor frequency estimator is derived. From this expression, an inequality describing the variance threshold of the estimator is found. The window weights are then optimized to improve the variance. Numerical simulations demonstrate that the variance approximations are valid for medium to high signal-to-noise ratios or for large numbers of samples.**

I. Introduction

To achieve the best possible use of processor time, computational efficiency in frequency estimation has become extremely important. Computationally efficient frequency estimators are those which can accurately estimate a single frequency in noise in $O(N)$ steps, given N samples. Several such algorithms have been proposed [1]–[4]. Of interest here is the weighted linear predictor proposed by Kay [2] because of its accuracy and simplicity.

In this paper, an approximate expression for the variance of Kay's weighted linear predictor is derived and compared with experimental results. It is shown that the derived expression is a good approximation at medium to high signal-to-noise ratios or for large numbers of samples.

It is shown that the weights can be optimized to minimize variance. The form of the resultant estimator is found to depend upon the signal-to-noise ratio as well as the number of points. Through examination of the optimal weights, a relationship is discovered between the Lank–Reed–Pollon estimator [1] and Kay's estimator. It is found that the maximum variance advantage of the optimal weights to Kay's weights is 20%.

II. Signal Model

Consider a complex sinusoidal signal which has been corrupted by noise in the receiver. The received samples s_n can be expressed as

$$s_n = Ae^{j(\omega n+\phi_0)} + z_n \tag{1}$$

Manuscript received March 2, 1993; revised December 21, 1993. This research was partially supported by the Cooperative Research Centre for Robust and Adaptive Systems ($(\mathrm{CR})^2$ASys), funded by the Australian Commonwealth Government under the Cooperative Research Centres Program. The associate editor coordinating the review of this paper and approving it for publication was Prof. John Goutsias.

V. Clarkson is with the Electronic Warfare Division, Defence Science and Technology Organisation, Salisbury, Australia.

P. J. Kootsookos is with the Department of Engineering, Australian National University, Canberra, Australia.

B. G. Quinn is with the Maritime Operations Division, Defence Science and Technology Organisation, Salisbury, Australia.

IEEE Log Number 9403257.

where $n = 0\ldots, N-1$, and N is the number of samples received, A is the amplitude of the sinusoid, ω is its frequency, ϕ_0 is the initial phase, and the z_n are the noise. The parameters A, ω, and ϕ_0 are unknown, and ω is to be estimated. It is assumed that the signal has been sampled in accordance with the Nyquist criterion for a unit time step. Hence, the frequency ω lies within the range $[-\pi, \pi)$.

The z_n are independent, identically distributed, zero-mean, complex normal random variables, with uncorrelated real and imaginary parts, each with variance σ^2. The signal-to-noise ratio (SNR) r can then be defined as $r = A^2/(2\sigma^2)$. The Cramér–Rao lower bound for the variance of an unbiased estimator of ω in (1) has been derived [5]. The lower bound is $6/\left[rN(N^2-1)\right]$.

Except where indicated, bias and variance are computed modulo 2π to account for the circular nature of the frequency ω. Hence, the bias of a frequency estimator is $\mathrm{E}[[\hat{\omega}-\omega]_{2\pi}]$ and variance of a frequency estimator is $\mathrm{E}\left[\{[\hat{\omega}-\omega]_{2\pi}\}^2\right]$, where $[x]_{2\pi} \equiv \{(x+\pi) \bmod (2\pi)\} - \pi$, *i.e.* $[x]_{2\pi} \in [-\pi, \pi)$.

The use of modulo statistics is preferred because they more accurately reflect the uncertainty of estimates of frequency. When an estimator is considered to be unbiased (modulo 2π), then such an estimator is not not necessarily unbiased in the normal, linear sense. To see this, consider a frequency estimate of a critically sampled signal with $\omega = -\pi$. Most estimators (and all of those considered here) will produce an estimate in the interval $[-\pi, \pi)$. Clearly, unless the estimator is completely without error, the estimator would be considered biased using linear statistics. However, sampled sinusoids with frequencies $-\pi + \delta$ and $\pi - \delta$ are practically indistinguishable when δ is sufficiently small. A good estimator should therefore return an estimate of the frequency $-\pi$, which is within a small interval greater than $-\pi$ or less than π. The use of modulo statistics takes this circular nature of frequency into account. A relationship between the linear and modulo bias and variance for Kay's weighted linear predictor is presented in Theorem 3, when it is known that $|\omega|$ is bounded away from π.

III. Kay's Weighted Linear Predictor

Kay [2] proposes a weighted linear predictor as an inferior alternative to his weighted phase averager (Kay's window estimator). The weighted linear predictor is given by

$$\hat{\omega}_{KC} = \angle \sum_{n=1}^{N-1} w_n s_n s_{n-1}^{\star} \tag{2}$$

1053-587X/94$04.00 © 1994 IEEE

where $w_n = 6n(N-n)/[N(N^2-1)]$ are the weights of Kay's window and $\angle$ denotes the complex argument.

IV. Approximation Of Variance

To analyze the variance of Kay's weighted linear predictor, the method of Lank *et al.* [1] is followed. First, (1) is rewritten as

$$s_n = (A+u_n)e^{j(\omega n+\phi_0)} \tag{3}$$

where now, u_n is a complex zero-mean random variable with $z_n = u_n e^{j(\omega n+\phi_0)}$, and the u_n have the same distribution as the z_n. It is then possible to rewrite (2) as

$$\begin{aligned}\hat{\omega}_{KC} &= \angle \sum_{n=1}^{N-1} w_n(A+u_n)(A+u_{n-1}^{\star})e^{j\omega} \\ &= \left[\omega + \angle \sum_{n=1}^{N-1} w_n\left(A^2 + Au_n + Au_{n-1}^{\star} + u_n u_{r-1}^{\star}\right)\right]_{2\pi}.\end{aligned} \tag{4}$$

Now, let $u_n = A(x_n + jy_n)$. Then, x_n and y_n are zero-mean independent, identically distributed, real Gaussian random variables, each with variance σ^2/A^2. Thus

$$\begin{aligned}&[\hat{\omega}_{KC} - \omega]_{2\pi} \\ &= \angle\left[\sum_{n=1}^{N-1} w_n(1 + x_n + x_{n-1} + x_n x_{n-1} + y_n y_{n-1}) \right. \\ &\quad \left. + j\sum_{n=1}^{N-1} w_n(x_{n-1}y_n - x_n y_{n-1} + y_n - y_{n-1})\right] \\ &= \angle(1 + X + jY)\end{aligned} \tag{5}$$

since $\sum_{n=1}^{N-1} w_n = 1$ and where

$$X = \sum_{n=1}^{N-1} w_n(x_n + x_{n-1} + x_n x_{n-1} + y_n y_{n-1}) \quad \text{and} \tag{6}$$

$$Y = \sum_{n=1}^{N-1} w_n(y_n - y_{n-1} + x_{n-1}y_n - x_n y_{n-1}). \tag{7}$$

Theorem 1: Kay's weighted linear predictor is an unbiased (modulo 2π) estimator of ω, i.e., $\mathrm{E}[[\hat{\omega}_{KC} - \omega]_{2\pi}] = 0$.

Proof: From (6) and (7), rewrite X and Y as $X(\boldsymbol{x},\boldsymbol{y})$, and $Y(\boldsymbol{x},\boldsymbol{y})$, respectively, where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the N-dimensional vectors formed from the x_n and y_n, respectively. Note that

$$\begin{aligned}X(\boldsymbol{x},-\boldsymbol{y}) &= X(\boldsymbol{x},\boldsymbol{y}), \\ Y(\boldsymbol{x},-\boldsymbol{y}) &= -Y(\boldsymbol{x},\boldsymbol{y}).\end{aligned}$$

Consider the N-dimensional vector $\boldsymbol{u}$ of $u_n = -y_n$, $n = 0, 1, \ldots, N-1$. Since the y_n are independent Gaussian random variables, the u_n have the same distribution as the y_n.

Now, consider the expected value of Kay's estimator. This can be written

$$\begin{aligned}\mathrm{E}[[\hat{\omega}_{KC} - \omega]_{2\pi}] &= \mathrm{E}[\angle\{1 + X(\boldsymbol{x},\boldsymbol{y}) + jY(\boldsymbol{x},\boldsymbol{y})\}] \\ &= \mathrm{E}[\angle\{1 + X(\boldsymbol{x},\boldsymbol{u}) + jY(\boldsymbol{x},\boldsymbol{u})\}] \\ &= \mathrm{E}[\angle\{1 + X(\boldsymbol{x},-\boldsymbol{y}) + jY(\boldsymbol{x},-\boldsymbol{y})\}] \\ &= \mathrm{E}[\angle\{1 + X(\boldsymbol{x},\boldsymbol{y}) - jY(\boldsymbol{x},\boldsymbol{y})\}] \\ &= -\mathrm{E}[\angle\{1 + X(\boldsymbol{x},\boldsymbol{y}) + jY(\boldsymbol{x},\boldsymbol{y})\}] \\ &= -\mathrm{E}[[\hat{\omega}_{KC} - \omega]_{2\pi}].\end{aligned}$$

Hence, $\mathrm{E}[[\hat{\omega}_{KC} - \omega]_{2\pi}] = 0$ and Kay's weighted linear predictor is unbiased, as required. □

The proof of the following theorem regarding the asymptotic variance of the estimator (modulo 2π) is to be found in the Appendix.

Theorem 2: The variance of $\hat{\omega}_{KC}$ (modulo 2π) is $\mathrm{E}[Y^2](1 + g_{N,r})$ where $g_{N,r} \longrightarrow 0$ as either $N \longrightarrow \infty$ or $r \longrightarrow \infty$, i.e.

$$\sigma^2_{\hat{\omega}_{KC}} = \frac{6}{rN(N^2-1)}\left\{1 + \frac{N^2+1}{10r}\right\}(1 + g_{N,r}). \tag{8}$$

Note that as $r \longrightarrow \infty$, $N^2/r \longrightarrow 0$, and so the ratio of (8) to the Cramér–Rao lower bound approaches unity. Clearly, Kay's estimator is not asymptotically statistically efficient, since $N^2/r \longrightarrow \infty$ as $N \longrightarrow \infty$.

From (8), the variance threshold can be calculated, i.e., that combination of N and r where the ratio of the variance of the estimator to the Cramér–Rao lower bound exceeds a given value. If this value is α, then the SNR boundary is given by

$$r > \frac{(N^2+1)}{10(\alpha-1)}. \tag{9}$$

Finally, if it is known *a priori* that $|\omega|$ is bounded away from π, then it becomes possible to obtain expressions for the linear bias and mean square error.

Theorem 3: Suppose that $|\omega| < \pi - \delta$, where $\delta > 0$. Then

$$\mathrm{E}[\hat{\omega}_{KC}] = \omega + \mathrm{o}(1) \qquad \text{and}$$
$$\mathrm{E}\left[(\hat{\omega}_{KC} - \omega)^2\right] = \sigma^2_{\hat{\omega}_{KC}}[1 + \mathrm{o}(1)]$$

as $N \longrightarrow \infty$ or $r \longrightarrow \infty$, where $\sigma^2_{\hat{\omega}_{KC}}$ is given by (8).

Proof: From (4) and (5), and letting $\theta_n = [\hat{\omega}_{KC} - \omega]_{2\pi} = \angle(1 + X + jY)$, the estimation error can be written as

$$\hat{\omega}_{KC} - \omega = \theta_n + 2\pi k_n$$

where

$$k_n = \begin{cases} -1 & \text{if } \theta_n + \omega \geqslant \pi, \\ 1 & \text{if } \theta_n + \omega < -\pi, \\ 0 & \text{otherwise.}\end{cases}$$

Since $\mathrm{E}[\theta_n] = 0$ from Theorem 1, it follows that $\mathrm{E}[\hat{\omega}_{KC} - \omega] = 2\pi\mathrm{E}[k_n]$. Now

$$\mathrm{E}[k_n] = \Pr[\theta_n + \omega < -\pi] - \Pr[\theta_n + \omega \geqslant \pi]$$

and therefore

$$\begin{aligned}|\mathrm{E}[k_n]| &\leqslant \Pr[\theta_n + \omega < -\pi] + \Pr[\theta_n + \omega \geqslant \pi] \\ &\leqslant \Pr[\theta_n < -\delta] + \Pr[\theta_n > \delta] \\ &= \Pr[|\theta_n| > \delta] \\ &\leqslant \delta^{-2K}\mathrm{E}\left[\theta_n^{2K}\right]\end{aligned}$$

by Markov's inequality, for any $K > 0$. Hence, $\hat{\omega}_{KC}$ is asymptotically unbiased (i.e., as $N \longrightarrow \infty$ or $r \longrightarrow \infty$) as $\mathrm{E}\left[\theta_n^{2K}\right] = \mathrm{o}(1)$, which can be easily shown using a similar method to that used in the proof of Theorem 2. Similarly

$$\begin{aligned}\mathrm{E}\left[(\hat{\omega}_{KC} - \omega)^2\right] &= \mathrm{E}\left[\theta_n^2\right] + 2\pi\mathrm{E}[\theta_n k_n] + 4\pi^2\mathrm{E}\left[k_n^2\right] \\ &= \mathrm{E}\left[\theta_n^2\right](1 + 2\pi\xi + 4\pi^2\zeta)\end{aligned}$$

where $\xi = \mathrm{E}[\theta_n k_n]/\mathrm{E}\left[\theta_n^2\right]$ and $\zeta = \mathrm{E}\left[k_n^2\right]/\mathrm{E}\left[\theta_n^2\right]$. Now $|\xi| \leqslant \sqrt{\zeta}$ by the Cauchy–Schwartz inequality and

$$\begin{aligned}\mathrm{E}\left[k_n^2\right] &= \Pr[\theta_n + \omega < -\pi] + \Pr[\theta_n + \omega \geqslant \pi] \\ &\leqslant \delta^{-2K}\mathrm{E}\left[\theta_n^{2K}\right]\end{aligned}$$

for any $K > 0$, as before. As the right hand side of the above is easily shown using the same techniques as in the proof of Theorem 2 to be $\mathrm{O}\left(r^{-K}N^{-K}\right)$, it follows that $\zeta = \mathrm{o}(1)$. Thus, $\mathrm{E}\left[(\hat{\omega}_{KC} - \omega)^2\right] = \mathrm{E}\left[\theta_n^2\right]\{1 + \mathrm{o}(1)\}$, as required. □

Note that in practice, the sizes of the $\mathrm{o}(1)$ terms in Theorem 3 are proportional to the size of $|\omega|$.

V. Derivation of the Optimal Window Weights

The optimal window weights can be derived for the linear predictor frequency estimator by considering the variance expression for a general weighting function and minimizing the variance over the window weights. Hence, the generalized linear predictor considered is

$$\hat{\omega}_{KC(opt)} = \angle \sum_{n=1}^{N-1} W_n x_n x_{n-1}^{\star} \tag{10}$$

where W_n are the optimal window weights to be determined.

Analysis of the variance of this estimator proceeds in a similar fashion to the previous analysis, but (26) must be rewritten, substituting the w_n with the W_n. The expression is then

$$\begin{aligned}\mathrm{E}\left[Y^2\right] &= W_1^2\mathrm{E}\left[y_0^2\right] + W_{N-1}^2\mathrm{E}\left[y_{N-1}^2\right] \\ &\quad + \sum_{n=2}^{N-1}(W_n - W_{n-1})^2\mathrm{E}\left[y_{n-1}^2\right] \\ &\quad + \sum_{n=1}^{N-1} W_n^2\left(\mathrm{E}\left[x_{n-1}^2 y_n^2\right] + \mathrm{E}\left[x_n^2 y_{n-1}^2\right]\right).\end{aligned}$$

By following the same method as used in Section IV, the variance of the optimal estimator can be written

$$\begin{aligned}\sigma^2_{\hat{\omega}_{KC(opt)}} &\approx \mathrm{E}\left[Y^2\right] \\ &= \frac{1}{r}\left(\sum_{n=1}^{N-1} W_n^2 + \sum_{n=2}^{N-1} W_n W_{n-1}\right) + \frac{1}{2r^2}\sum_{n=1}^{N-1} W_n^2.\end{aligned}$$

To derive the optimal window weights, it is necessary to minimize the variance with respect to the weights, with the condition that $\sum_{n=1}^{N-1} W_n = 1$, and this condition can be satisfied by using the method of Lagrange multipliers. We thus minimize

$$C = \sigma^2_{\hat{\omega}_{KC(opt)}} + \lambda\left(\sum_{n=1}^{N-1} W_n - 1\right)$$

with respect to the W_n. The set of simultaneous equations to be solved is

$$\frac{\partial C}{\partial W_i} = \frac{2W_i}{r} - \frac{W_{i-1}}{r} - \frac{W_{i+1}}{r} + \frac{W_i}{r^2} + \lambda = 0$$

for $i = 1, \ldots, N-1$. These equations can be rewritten as the difference equation

$$W_i - \left(2 + \frac{1}{r}\right)W_{i-1} + W_{i-2} = \lambda r \tag{11}$$

with the conditions

$$W_0 = 0 \tag{12}$$

$$W_N = 0 \tag{13}$$

$$\sum_{n=1}^{N-1} W_n = 1. \tag{14}$$

Now, the solution of the homogeneous equation is

$$W_n = a\beta^n + b\beta^{-n}$$

where a and b are arbitrary constants depending on the initial conditions and β is a root of the auxiliary equation. Hence

$$\beta = 1 + \frac{1}{2r} + \sqrt{\frac{1}{4r^2} + \frac{1}{r}}.$$

The solution to (11) is thus

$$W_n = a\beta^n + b\beta^{-n} + c$$

where a, b and c must satisfy the conditions of (12)–(14), i.e., they satisfy

$$a + b + c = 0, \tag{15}$$

$$a\beta^N + b\beta^{-N} + c = 0, \tag{16}$$

$$a\sum_{n=1}^{N-1}\beta^n + b\sum_{n=1}^{N-1}\beta^{-n} + c(N-1) = 1. \tag{17}$$

Thus

$$\begin{bmatrix} 1 & 1 & 1 \\ \beta^N & \beta^{-N} & 1 \\ \frac{\beta^N - \beta}{\beta - 1} & \frac{1 - \beta^{-N+1}}{\beta - 1} & N-1 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{18}$$

The solution to these simultaneous linear equations is

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \Delta^{-1}\begin{bmatrix} \beta^{-N} - 1 \\ 1 - \beta^N \\ \beta^N - \beta^{-N} \end{bmatrix}$$

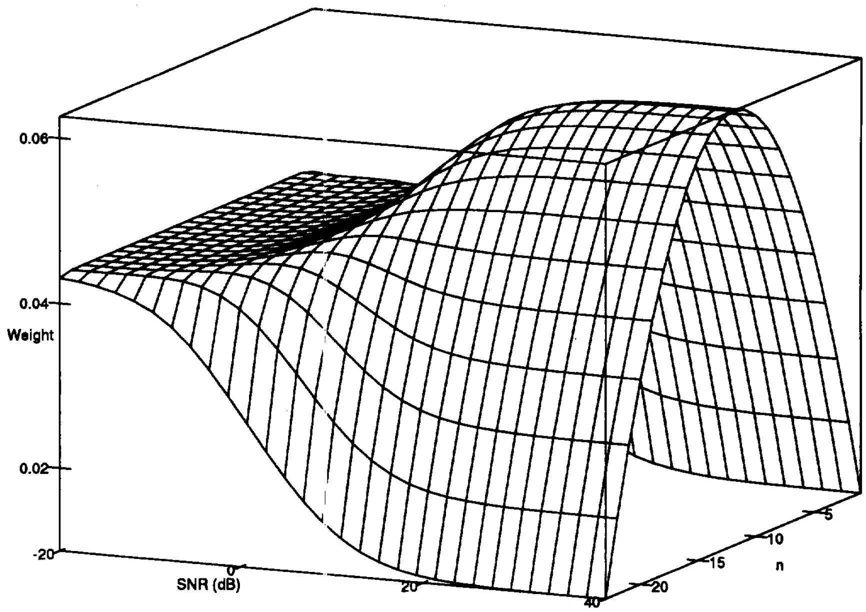


Fig. 1. Optimal window weights for the weighted linear predictor when $N = 24$.

where Δ is the determinant of the matrix of coefficients in (18). Hence

$$\begin{aligned} W_n &= \Delta^{-1}\left[\beta^N - \beta^{-N} + (\beta^{-N} - 1)\beta^n + (1-\beta^N)\beta^{-n}\right] \\ &= \Delta^{-1}\left[(\beta^N - \beta^{-N}) - (\beta^n - \beta^{-n}) - (\beta^{N-n} - \beta^{n-N})\right] \\ &= 2\Delta^{-1}[\sinh(N\theta) - \sinh(n\theta) - \sinh\{(N-n)\theta\}] \end{aligned} \tag{19}$$

where $\theta = \ln \beta$ and

$$\begin{aligned} \frac{\Delta}{2} &= \sum_{n=1}^{N-1} \sinh(N\theta) - \sinh(n\theta) - \sinh\{(N-n)\theta\} \\ &= (N-1)\sinh(N\theta) \\ &\quad - 2\sinh\left(\frac{1}{2}N\theta\right)\sinh\left\{\frac{1}{2}(N-1)\theta\right\}/\sinh\left(\frac{1}{2}\theta\right). \end{aligned}$$

Hence, the optimal window weights for the linear predictor frequency estimator can be written as (20), which appears at the bottom of the page. Note that the normalizing constraint of (17) can be relaxed, and the $2\Delta^{-1}$ term removed.

Interestingly, this equation depends not only on the number of points, N, but also on the signal-to-noise ratio, r, since θ is a function of r. In practice, this will mean that the optimal window may be difficult to use since the SNR is seldom known *a priori*. The shape of the optimal window is presented for a range of SNR's in Fig. 1. It is now shown that in the limits as $N \longrightarrow \infty$ and $r \longrightarrow 0$ the optimal window converges to a rectangular window, and that as $r \longrightarrow \infty$, the optimal window converges to Kay's window.

Theorem 4: The following limits apply for the optimal window weights W_n of the weighted linear predictor frequency estimator:

$$\lim_{r \longrightarrow 0} (N-1)W_n = 1 \tag{21}$$

$$\lim_{r \longrightarrow \infty} \frac{W_n}{w_n} = 1. \tag{22}$$

Proof of (21): In the limit as $r \longrightarrow 0$, $\beta \longrightarrow \infty$. Hence, $\sinh(k\theta) = \frac{1}{2}\beta^k + O(\beta^{-k})$ and so

$$\begin{aligned} \lim_{r \longrightarrow 0} (N-1)W_n &= \lim_{\beta \longrightarrow \infty} (N-1)W_n \\ &= \lim_{\beta \longrightarrow \infty} (N-1)\frac{\frac{1}{2}\beta^N + O(\beta^{N-1})}{(N-1)\frac{1}{2}\beta^N + O(\beta^{N-1})} \\ &= 1 \end{aligned}$$

as required. □

Proof of (22): In the limit as $r \longrightarrow \infty$, $\theta \longrightarrow 0$. From the Taylor's series expansion, $\sinh(k\theta) = k\theta + k^3\theta^3/6 + O(\theta^4)$ for any real k, and hence

$$\lim_{r \longrightarrow \infty} \frac{W_n}{w_n} = \lim_{\theta \longrightarrow 0} \frac{W_n}{w_n}.$$

$$W_n = \frac{\sinh(N\theta) - \sinh(n\theta) - \sinh\{(N-n)\theta\}}{(N-1)\sinh(N\theta) - 2\sinh\left(\frac{1}{2}N\theta\right)\sinh\left\{\frac{1}{2}(N-1)\theta\right\}/\sinh\left(\frac{1}{2}\theta\right)}. \tag{20}$$

Consider the numerator of W_n in (20). Now

$$\begin{aligned}\sinh(N\theta) - \sinh(n\theta) - \sinh\{(N-n)\theta\} &= N\theta + \frac{N^3\theta^3}{8} - n\theta - \frac{n^3\theta^3}{8} - (N-n)\theta \\ &\quad - \frac{(N-n)^3\theta^3}{8} + \mathrm{O}(\theta^4) \\ &= \frac{1}{2}Nn(N-n)\theta^3 + \mathrm{O}(\theta^4).\end{aligned}$$

Similarly, the denominator of W_n can be written

$$\begin{aligned}&(N-1)\sinh(N\theta) - 2\sinh\left(\frac{1}{2}N\theta\right) \\ &\qquad \sinh\left\{\frac{1}{2}(N-1)\theta\right\}/\sinh\left(\frac{1}{2}\theta\right) \\ &\quad = (N-1)\sinh(N\theta) - 2\sum_{n=1}^{N-1}\sinh(n\theta)\end{aligned}$$

$$\begin{aligned}&= (N-1)\sinh(N\theta) - 2\sum_{n=1}^{N-1}\sinh(n\theta) \\ &= (N-1)\left(N\theta + \frac{N^3\theta^3}{6}\right) \\ &\quad - 2\sum_{n=1}^{N-1}\left(n\theta + \frac{n^3\theta^3}{6}\right) + \mathrm{O}(\theta^4) \\ &= N(N-1)\theta + \frac{N^3(N-1)\theta^3}{6} \\ &\quad - 2\left\{\frac{N(N-1)\theta}{2} + \frac{N^2(N-1)^2\theta^3}{24}\right\} + \mathrm{O}(\theta^4) \\ &= \frac{N^2(N^2-1)\theta^3}{12} + \mathrm{O}(\theta^4).\end{aligned}$$

Hence

$$\begin{aligned}\frac{W_n}{w_n} &= \frac{1}{w_n}\frac{\frac{1}{2}Nn(N-n)\theta^3 + \mathrm{O}(\theta^4)}{\frac{1}{12}N^2(N^2-1)\theta^3 + \mathrm{O}(\theta^4)} \\ &= \frac{1}{w_n}\left\{\frac{6n(N-n)}{N(N^2-1)} + \mathrm{O}(\theta)\right\} \\ &= \frac{1}{w_n}\{w_n + \mathrm{O}(\theta)\}\end{aligned}$$

and so, $\lim_{\theta \longrightarrow 0} W_n/w_n = 1$, as required. □

The theorem shows that there is no advantage (and in fact a disadvantage) in using Kay's window for low SNR. It is also true that, for a given SNR, the shape of the optimal window approaches the rectangular window as $N \longrightarrow \infty$.

From [1], an approximation for the variance of the first-difference Lank–Reed–Pollon estimator (which is equivalent to the weighted linear predictor frequency estimator with a rectangular window) is

$$\sigma^2_{\hat{\omega}_{LRP}} = \frac{2r + N - 1}{2(N-1)^2 r^2}.$$

Therefore, the limit of the ratio of the variance of the Kay's window estimator to the optimal window estimator variance as $N \longrightarrow \infty$ will be the same as the limit of the ratio of the variances of the Kay's window estimator to the Lank–Reed–Pollon estimator. Hence

$$\begin{aligned}\lim_{N\longrightarrow\infty}\frac{\sigma^2_{\hat{\omega}_{KC}}}{\sigma^2_{\hat{\omega}_{KC(opt)}}} &= \lim_{N\longrightarrow\infty}\frac{\sigma^2_{\hat{\omega}_{KC}}}{\sigma^2_{\hat{\omega}_{LRP}}} \\ &= \lim_{N\longrightarrow\infty}\frac{6}{rN(N^2-1)}\left(1 + \frac{N^2+1}{10r}\right) \\ &\quad \cdot \frac{2(N-1)^2 r^2}{2r + N - 1} \\ &= \frac{6}{5}.\end{aligned} \tag{23}$$

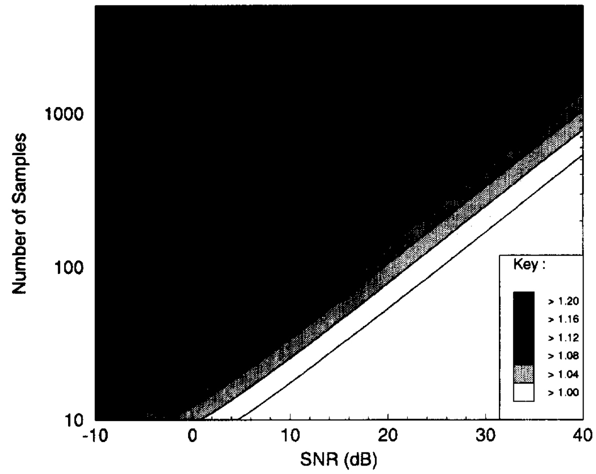


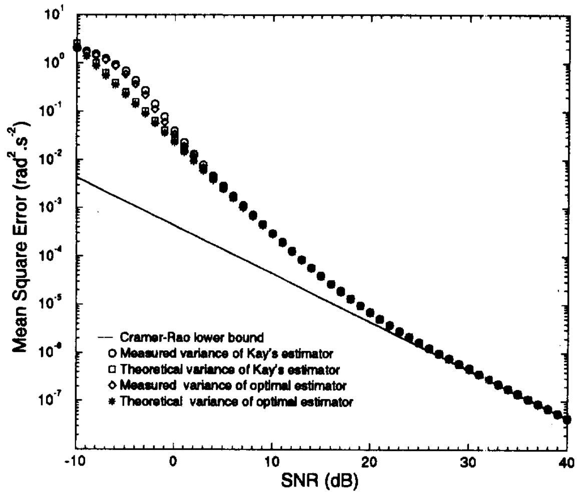
Fig. 2. Ratio of the variances of the weighted linear predictor frequency estimator using Kay's window against using the optimal window.

Therefore, when N is large, the advantage conferred by using the optimal window over the Kay's window is an improvement in the variance of 20%. The ratios of variances of Kay's window to the optimal window are presented for a range of N's and SNR's in Fig. 2. The plot shows that the improvement gained by using the optimal window is always less than 20%, as predicted by the theory. These results suggest that there is little to be gained from using the optimal window. However, it also suggests that in low SNR environments or where there is a large number of samples, the best computational solution may be the Lank–Reed–Pollon (i.e., rectangular window) estimator.

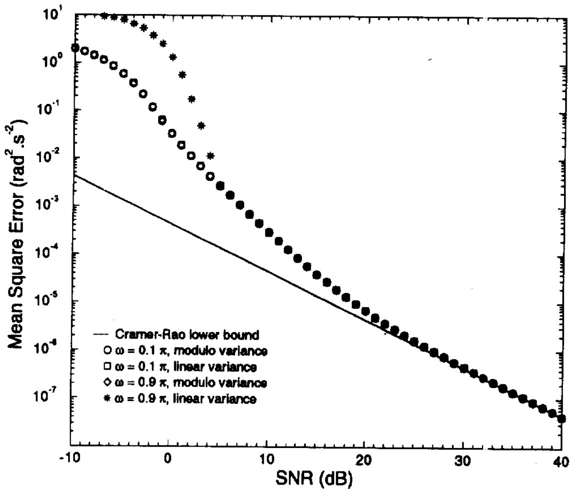
VI. Simulations

To test the validity of the approximation for the variance of Kay's weighted linear predictor of (8), a large number of simulations were performed on a MASPAR MP-1 massively parallel computer.

Mean square errors were estimated from the average of the square errors of each of 20 480 trial runs performed for the estimator for a given r and N. Each set of observations contains a signal of frequency ω and initial phase ϕ_0 both uniformly distributed in $[-\pi, \pi)$ so that the behavior of the estimators can be ascertained when the signal frequency is unknown.



(a)



(b)

Fig. 3 Comparison of estimators' statistical performance and Cramér–Rao lower bound with SNR for N = 24: (a) Estimators' theoretical and measured variance; (b) optimal estimator's modulo and linear measured variance for two frequencies.

For the case of $N = 24$ points, the mean square errors from the simulations are compared with the theoretical variances and the Cramér–Rao lower bound in Fig. 3(a). It can be seen that the theory is in close agreement with the simulations when the SNR is greater than 5 dB. However, a plot for fixed N does not give a picture of how well the theory predicts variance over a range of Ns. It can also be seen that there is very little difference between the performance of Kay's estimator and the optimally weighted linear predictor.

A comparison of modulo and linear statistics is presented in Fig. 3(b) for the optimal estimator for $N = 24$ points for selected frequencies. Clearly, for variance calculated modulo 2π, there is no discernible difference between the measured variance for $\omega = 0.1\pi$ and the measured variance for $\omega = 0.9\pi$, as is to be expected from the theory. However, when using linear statistics, there is a clear difference in the performance at low SNR, with the measured variance being significantly larger for $\omega = 0.9\pi$. Again, this has been predicted by the theory.

In Fig. 4(a), measured mean square errors for a range of SNR's and N's are presented as a contour plot. The contours are those of equal ratio of the measured mean square error to the Cramér–Rao lower bound. The white region indicates where the ratio is less than 5 dB. In Fig. 4(b), a contour plot has been generated from the theoretical approximation to the variance of (8). From the plots, it is clear that the theoretical approximation to the variance appears to be a good one for medium to high SNR, with significant differences appearing only when the product rN is small.

It is clear that the measured and theoretical ratios of the optimally weighted linear predictor in Fig. 5 are nearly identical to those seen in Fig. 4 for Kay's weighted linear predictor. The slight improvement in variance of up to 20%, as predicted in (23), is too small to be clearly visible in the contour plots.

VII. Conclusion

A theoretical approximation has been derived for the variance of Kay's weighted linear predictor. A simple expression to describe the variance threshold of the estimator was also derived. The weights were then optimized to minimize variance. The optimal weights depend on the signal-to-noise ratio as well as the number of samples. The variance advantage gained was found to be less than 20%, but it was found that a simple unweighted linear predictor (or Lank–Reed–Pollon) frequency estimator may be a superior solution (in terms of variance and computational load) when the signal-to-noise ratio is low or the number of samples is high. It was demonstrated by extensive numerical simulations that the variance approximations were valid whenever r (or N) were large.

Appendix
Proofs of Theorems and Lemmata

Lemma 1: The even moments of X and Y, $\mathrm{E}[X^{2k}]$ and $\mathrm{E}[Y^{2k}]$, as defined in (6) and (7), are both $\mathrm{O}(N^{-k}r^{-k})$.

Proof: From (6)

$$\begin{aligned} X &= \sum_{n=1}^{N-1} w_n(x_n + x_{n-1} + x_n x_{n-1} + y_n y_{n-1}) \\ &= \underbrace{\sum_{n=0}^{N-1} (w_n + w_{n+1})x_n}_{Q} + \underbrace{\sum_{n=1}^{N-1} w_n(x_n x_{n-1} + y_n y_{n-1})}_{R}. \end{aligned}$$

It follows that

$$\begin{aligned} \mathrm{E}[X^{2k}] &= \sum_{j=0}^{2k} \binom{2k}{j} \mathrm{E}[Q^j R^{2k-j}] \\ &\leqslant \sum_{j=0}^{2k} \binom{2k}{j} \sqrt{\mathrm{E}[Q^{2j}]\mathrm{E}[R^{4k-2j}]}. \end{aligned}$$

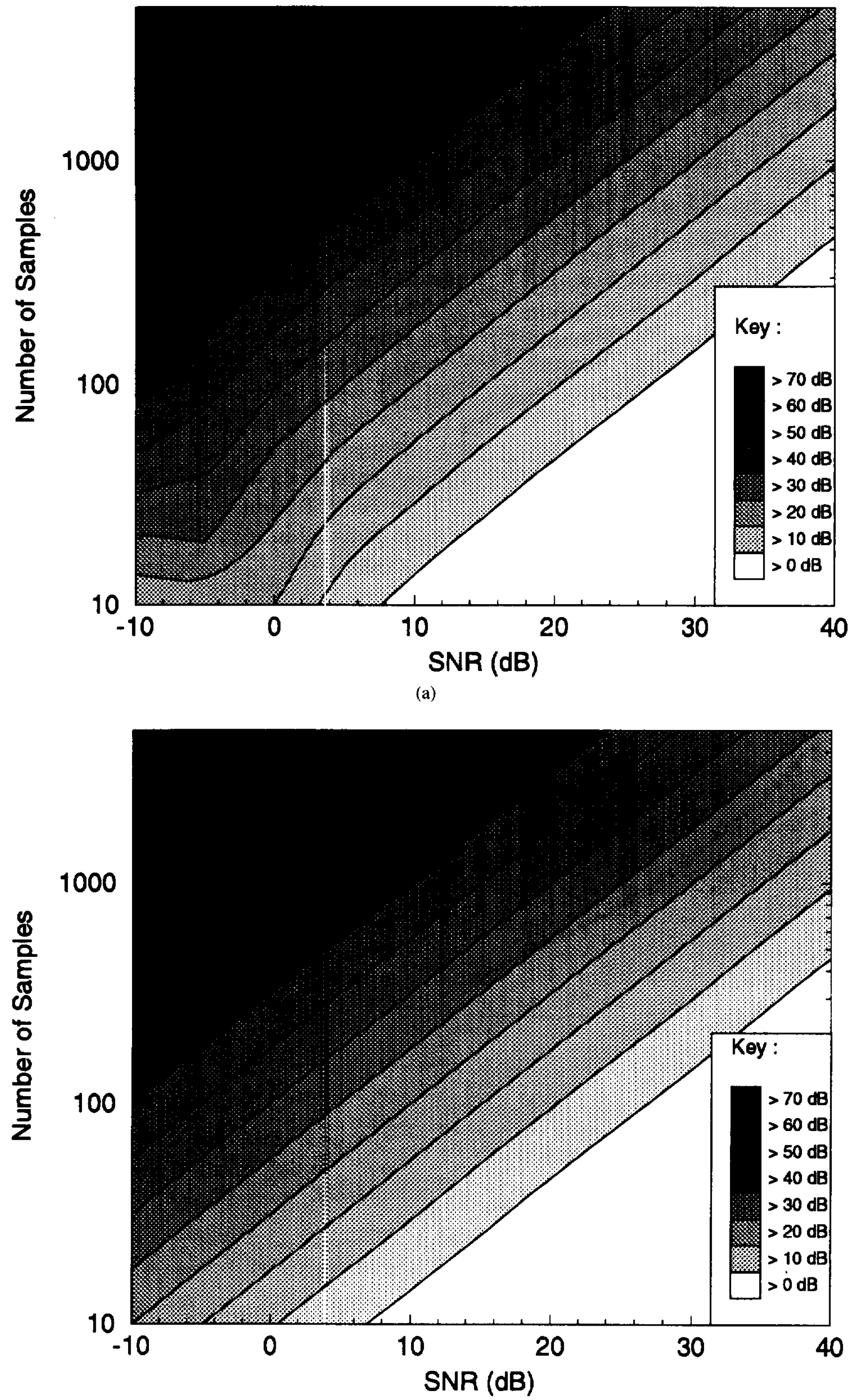


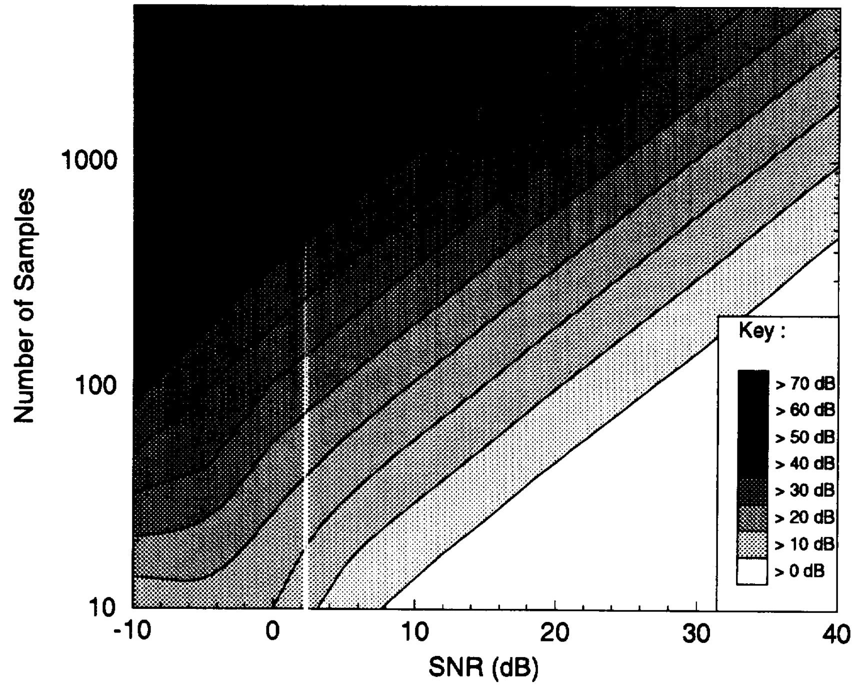
Fig. 4. Ratios of mean square error to Cramér–Rao lower bound in decibels for Kay's weighted linear predictor: (a) Measured ratio; (b) theoretical ratio.

As $(w_n + w_{n+1})$ is $\mathrm{O}(N^{-1})$ for each n, it follows that Q is normally distributed with mean zero and variance $\mathrm{O}(r^{-1}N^{-1})$. Consequently, $\mathrm{E}[Q^{2j}]$ is $\mathrm{O}(r^{-j}N^{-j})$. In a similar way, it may be shown that $\mathrm{E}[R^{2j}]$ is $\mathrm{O}(r^{-2j}N^{-j})$.

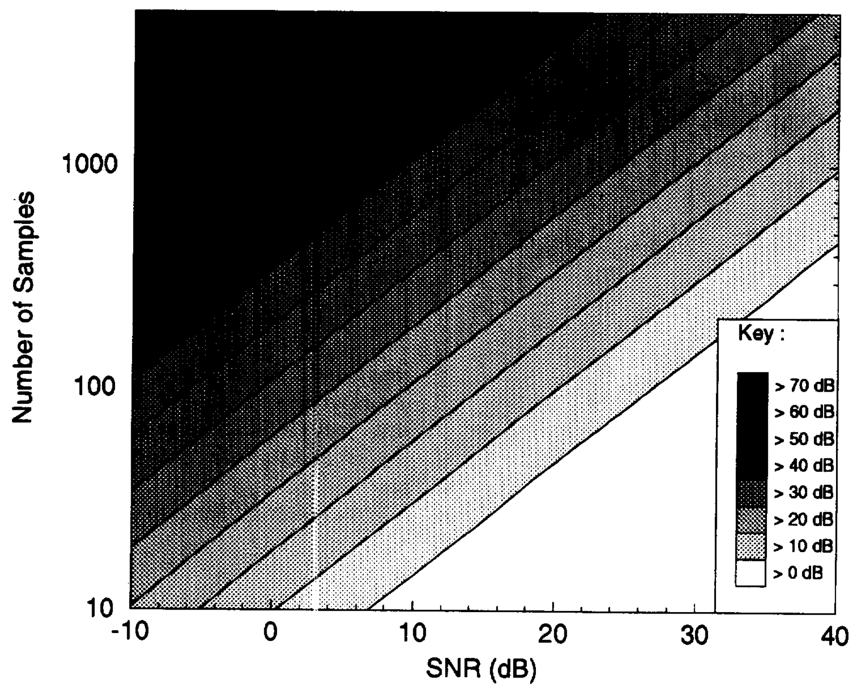
Thus

$$\mathrm{E}[X^{2k}] \leqslant \sum_{j=0}^{2k} \binom{2k}{j} \mathrm{O}\left(r^{j/2-2k}N^{-k}\right) = \mathrm{O}(N^{-k}r^{-k}).$$

Now, from (7)



(a)



(b)

Fig. 5. Ratios of mean square error to Cramér–Rao lower bound in decibels for the optimal weighted linear predictor: (a) Measured ratio; (b) theoretical ratio.

$$Y = \underbrace{\sum_{n=0}^{N-1} (w_n - w_{n+1}) y_n}_{S} + \underbrace{\sum_{n=1}^{N-1} w_n (x_{n-1} y_n - x_n y_{n-1})}_{T}.$$

As $(w_n - w_{n+1})$ is $\mathrm{O}(N^{-2})$, the term S is normally distributed with mean zero and variance $\mathrm{O}(r^{-1}N^{-3})$. Thus, $\mathrm{E}[S^{2j}]$ is $\mathrm{O}(r^{-j}N^{-3j})$. It is also easily shown that $\mathrm{E}[T^{2j}]$ is $\mathrm{O}(r^{-2j}N^{-j})$. Thus

$$\begin{aligned}\mathrm{E}[Y^{2k}] &= \sum_{j=0}^{2k}\binom{2k}{j}\mathrm{E}[S^jT^{2k-j}]\\ &\leqslant \sum_{j=0}^{2k}\binom{2k}{j}\mathrm{O}\left(r^{j/2-2k}N^{-j-k}\right)\\ &= \mathrm{O}(N^{-k}r^{-k}).\end{aligned}$$

□

Lemma 2: The variance of Y is

$$\mathrm{E}[Y^2] = \frac{6}{rN(N^2-1)}\left\{1+\frac{N^2+1}{10r}\right\}.$$

Proof: Rewriting Y thus

$$\begin{aligned}Y &= \frac{6}{N(N^2-1)}\sum_{n=1}^{N-1} n(N-n)(y_n - y_{n-1} + x_{n-1}y_n - x_n y_{n-1})\\ &= \frac{6}{N(N^2-1)}\left\{\sum_{n=0}^{N-1}(2n+1-N)y_n\right.\\ &\quad \left. + \sum_{n=1}^{N-1} n(N-n)(x_{n-1}y_n - x_n y_{n-1})\right\} \end{aligned} \tag{24}$$

the variance of Y, $\mathrm{E}[Y^2]$, can be found, and, ignoring the cross-products whose expectations are zero, this can be expressed as

$$\begin{aligned}\mathrm{E}[Y^2] &= \frac{36}{N^2(N^2-1)}\left\{\sum_{n=0}^{N-1}(2n+1-N)^2\mathrm{E}[y_n^2]\right.\\ &\quad \left. + \sum_{n=1}^{N-1} n^2(n-N)^2\left(\mathrm{E}[x_{n-1}^2y_n^2] + \mathrm{E}[x_n^2y_{n-1}^2]\right)\right\}\\ &= \frac{36}{N^2(N^2-1)}\left\{\frac{N}{6r}+\frac{N(N^2+1)}{60r^2}\right\}\\ &= \frac{6}{rN(N^2-1)}\left\{1+\frac{N^2+1}{10r}\right\}. \end{aligned} \tag{25}$$

□

Lemma 3: The moments of X^2+Y^2, i.e., $\mathrm{E}\left[(X^2+Y^2)^k\right]$, are $\mathrm{O}(N^{-k}r^{-k})$.

Proof:

$$\begin{aligned}\mathrm{E}\left[(X^2+Y^2)^k\right] &= \sum_{j=0}^{k}\binom{k}{j}\mathrm{E}[X^{2j}Y^{2k-2j}]\\ &\leqslant \sum_{j=0}^{k}\binom{k}{j}\sqrt{\mathrm{E}[X^{4j}]\mathrm{E}[Y^{4k-4j}]}.\end{aligned}$$

From Lemma 1, this then becomes

$$\begin{aligned}[E(X^2+Y^2)^k] &= \sum_{j=0}^{k}\binom{k}{j}\sqrt{\mathrm{O}(N^{-2j}r^{-2j})\mathrm{O}(N^{2j-2k}r^{2j-2k})}\\ &= \sum_{j=0}^{k}\binom{k}{j}\sqrt{\mathrm{O}(N^{-2k}r^{-2k})}\\ &= \sum_{j=0}^{k}\binom{k}{j}\mathrm{O}(N^{-k}r^{-k})\\ &= \mathrm{O}(N^{-k}r^{-k}).\end{aligned}$$

□

Proof of Theorem 2: Let $a(X,Y) = [\hat{\omega}_{KC}-\omega]_{2\pi} = \angle(1+X+jY)$ from (5). Then $\mathrm{E}[a(X,Y)] = 0$ from Theorem 1 and thus

$$\begin{aligned}\mathrm{E}\left[\{[\hat{\omega}_{KC}-\omega]_{2\pi}\}^2\right] &= \underbrace{\int_{x^2+y^2\leqslant\epsilon^2} a^2(x,y)f(x,y)\,dx\,dy}_{U}\\ &\quad + \underbrace{\int_{x^2+y^2>\epsilon^2} a^2(x,y)f(x,y)\,dx\,dy}_{V}\end{aligned}$$

where $f(x,y)$ is the joint p.d.f. of X and Y. Now

$$\begin{aligned}V &\leqslant \pi^2\mathrm{Pr}[X^2+Y^2>\epsilon^2]\\ &\leqslant \pi^2\epsilon^{-2k}\mathrm{E}\left[(X^2+Y^2)^k\right]\end{aligned}$$

by Markov's inequality and, for $\epsilon < 1$

$$\begin{aligned}U &= \int_{x^2+y^2\leqslant\epsilon^2}\left[\tan^{-1}\{y/(1+x)\}\right]^2 f(x,y)\,dx\,dy\\ &= \int_{x^2+y^2\leqslant\epsilon^2}\{y+\mathrm{O}(\epsilon^2)\}^2 f(x,y)\,dx\,dy\\ &= \int_{x^2+y^2\leqslant\epsilon^2}\{y^2+\mathrm{O}(\epsilon^3)\}f(x,y)\,dx\,dy\\ &= \mathrm{E}[Y^2]+\mathrm{O}(\epsilon^3)-\mathrm{E}[Y^2I]\end{aligned} \tag{26}$$

where I is 1 when $X^2+Y^2>\epsilon^2$ and 0 otherwise, and the orders are taken as $\epsilon \longrightarrow 0$. Thus, it is required to show that ϵ and k may be chosen so that V and the rightmost two terms of (26) go to zero, when divided by $\mathrm{E}[Y^2]$, as either $N \longrightarrow \infty$ or $r \longrightarrow \infty$.

In the following, repeated use shall be made of the inequality

$$|\mathrm{E}[LM]| \leqslant \sqrt{\mathrm{E}[L^2]\mathrm{E}[M^2]}.$$

It is easily seen that $V/\mathrm{E}[Y^2] \longrightarrow 0$ as $N \longrightarrow \infty$ or $r \longrightarrow \infty$ if

$$\epsilon^{2k} = \kappa^{-1}\frac{\mathrm{E}\left[(X^2+Y^2)^k\right]}{\mathrm{E}[Y^2]} \tag{27}$$

where $\kappa \longrightarrow 0$ as either $r \longrightarrow \infty$ or $N \longrightarrow \infty$. From Lemma 2 and Lemma 3, (27) can be rewritten

$$\epsilon^{2k} = \kappa^{-1}\mathrm{O}(N^{1-k}r^{1-k}).$$

If $k \geqslant 2$, κ can be chosen so that this expression approaches zero as $N \longrightarrow \infty$ or $r \longrightarrow \infty$.

Furthermore,

$$\frac{\mathrm{E}[Y^2 I]}{\mathrm{E}[Y^2]} \leqslant \frac{\sqrt{\mathrm{E}[Y^4]\mathrm{E}[I^2]}}{\mathrm{E}[Y^2]} = \frac{\sqrt{\mathrm{E}[Y^4]}}{\mathrm{E}[Y^2]}\sqrt{\Pr[X^2+Y^2>\epsilon^2]}.$$

From Lemma 1 and Lemma 2 it is clear that

$$\frac{\sqrt{\mathrm{E}[Y^4]}}{\mathrm{E}[Y^2]} = \mathrm{O}(1)$$

as $N \longrightarrow \infty$ or $r \longrightarrow \infty$. Also, as $\mathrm{E}[Y^2] \longrightarrow 0$ and $V/\mathrm{E}[Y^2] \longrightarrow 0$ as $r \longrightarrow \infty$ or $N \longrightarrow \infty$, so does $\Pr[X^2+Y^2>\epsilon^2]$. Consequently, $\mathrm{E}[Y^2 I]/\mathrm{E}[Y^2] \longrightarrow 0$.

It remains to find k and κ such that $\epsilon^3/\mathrm{E}[Y^2] \longrightarrow 0$ or, equivalently, $\epsilon^{6k}/\mathrm{E}[Y^2]^{2k} \longrightarrow 0$. From Lemma 2,

$$\frac{\epsilon^{6k}}{\mathrm{E}[Y^2]^{2k}} = \kappa^{-3}\mathrm{O}\left(N^{3-k}r^{3-k}\right).$$

It is clear, therefore, that if $k \geqslant 4$, κ may be chosen in such a way that this expression, and consequently ϵ, approaches zero as either $N \longrightarrow \infty$ or $r \longrightarrow \infty$.

Hence, it has been shown that the variance of $\hat{\omega}_{KC}$ is $\mathrm{E}[Y^2](1+g_{N,r})$ where $g_{N,r} \longrightarrow 0$ as either $N \longrightarrow \infty$ or $r \longrightarrow \infty$, as required. □

ACKNOWLEDGMENT

The authors are grateful to the Centre for Information and Technology Research (CiTR), University of Queensland, for the generous donation of computer processor time.

REFERENCES

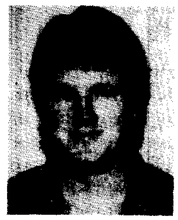
[1] G. W. Lank, I. S. Reed, and G. E. Pollon, "A semicoherent detection and doppler estimation statistic," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-9, pp. 151–165, Mar. 1973.

[2] S. M. Kay, "A fast and accurate single frequency estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1987–1990, Dec. 1989.

[3] B. C. Lovell and R. C. Williamson, "The statistical performance of some instantaneous frequency estimators," *IEEE Trans. Signal Processing*, vol. 40, pp. 1708–1723, July 1992.

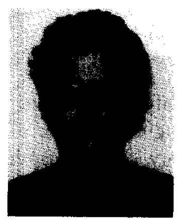
[4] V. Clarkson, "Efficient single frequency estimators," in *Proc. ISSPA*, Aug. 1992, pp. 327–330.

[5] D. C. Rife and R. R. Boorstyn, "Single-tone parameter estimation from discrete-time observations," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 591–598, Sept. 1974.



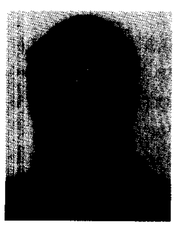
Vaughan Clarkson (S'93) was born in Brisbane, Queensland, Australia, in 1968. He received the B.Sc. degree in mathematics and the B.E. (Hons.) degree in computer systems engineering from the University of Queensland in 1989 and 1990. Currently, he is enrolled in a Ph.D. program with the Department of Systems Engineering at the Australian National University, Canberra. The area of his research is pulse train signal processing.

Since 1987, he has been employed by the Defence Science and Technology Organisation, Salisbury, initially as a cadet and since 1991 as an engineer. He is a member of the Cooperative Research Centre for Robust and Adaptive Systems.



Peter J. Kootsookos (S'86–M'92) received the B.E. and M.Eng.Sc. degrees in electrical engineering in 1986 and 1989, respectively, from the University of Queensland, Brisbane, and the Ph.D. degree in systems engineering from the Australian National University, Canberra, in 1992.

He is currently working as a Lecturer in the Department of Engineering at the Australian National University. His research interests include FIR filter bank design, robust control, spectrum estimation, and frequency estimation.



Barry G. Quinn (M'91) was born in Maitland, New South Wales, Australia, in 1956. He received the B.A. and Ph.D. degrees in statistics in 1977 and 1980 from the Australian National University.

He has held lectureships at the Universities of Wollongong (1981–1983) and Queensland (1983–1987) and a senior lectureship at the University of Newcastle (1987–1990). In 1990, he joined the Defence Science and Technology Organisation as a Principal Research Scientist. His research interests include the estimation and tracking of frequency and bearing and, more generally, the analysis of time series with quasi-sinusoidal components..